UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,174	10/31/2000	Charu C. Aggarwal	YOR920000430US1	7445

7590 07/18/2008

William E Lewis
Ryan Mason & Lewis LLP
90 Forest Avenue
Locust Valley, NY 11560

EXAMINER	
HILLERY, NATHAN	

ART UNIT	PAPER NUMBER
2176	

MAIL DATE	DELIVERY MODE
07/18/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Application Number: 09/703,174
Filing Date: October 31, 2000
Appellant(s): AGGARWAL ET AL.

David E. Shifren
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed 5/12/08 appealing from the Office action mailed 11/8/07.

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

No amendment after final has been filed.

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is substantially correct. The changes are as follows:

Claims 9, 18 and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chakrabarti et al. and Chaudhuri et al. (US 6529901 B1) as applied to claims 1 – 8, 10 – 17 and 19 – 26 above, and further in view of Chakrabarti et al. (Distributed Hypertext Resource Discovery Through Examples) [as cited by appellant] later referenced as Ch2 et al.

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

S. Chakrabarti et al., "Focused Crawling: A New Approach to Topic-Specific Web Resource Discovery," Computer Networks, 25 pages, 1999.

S. Chakrabarti et al., "Distributed Hypertext Resource Discovery Through Examples," Proceedings of the 25th VLDB Conference, Edinburgh, Scotland, pp. 375-386, 1999. (referenced as Ch2 et al.)

6529901 Chaudhuri et al. 3-2003

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1 – 8, 10 – 17 and 19 – 26 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chakrabarti et al. (Focused Crawling: A New Approach to Topic-specific Web Resource Discovery) [as cited by appellant] and in further view of Chaudhuri et al. (US 6529901 B1).

Regarding independent claim 1, Chakrabarti et al. teach that *keyword search is used to locate an initial set of pages (using a giant crawl and index)* (p 6, section 2.2, last paragraph), which meet the limitation of **initially retrieving one or more documents from the information network that satisfy a user-defined predicate,**

wherein the initial document retrieval operation is performed without assuming a specific model of a linkage structure such that the initial document retrieval operation retrieves the one or more documents without assuming that a relationship exists between a feature of a first one of the one or more documents and a feature of at least another one of the one or more documents that links to the first one.

Chakrabarti et al. teach that while fetching a document, the above formulation is used to find the leaf node with the highest probability. If some ancestor has been marked good we allow future visitation of URLs found on the document, otherwise the crawl is pruned there (p 9, section Hard focus rule), which meet the limitation of **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed and using the collected statistical information to automatically determine further document retrieval operations to be performed in accordance with the information network**, since the probabilities are calculated to find the "best" leaf node, the ancestors are analyzed to determine if they are good, and then based on that finding future visitations are allowed (p 9, section Hard focus rule). It should be noted that the *probabilities* of Chakrabarti et al. are equivalent to the claimed **statistical information.**

Chakrabarti et al. teach that a focused crawler is an example-driven automatic porthole-generator. We feel that the ability to focus on a topical subgraph of the Web, as in this paper, together with the ability to browse communities within that subgraph, will lead to significantly improved Web resource discovery (p 3, last paragraph before

Section 2), which meet the limitation of **wherein the statistical information-using step further comprises learning a linkage structure from at least a portion of the collected statistical information with each successive document retrieval operation such that the learned linkage structure is available for use in performing subsequent document retrieval operations requested by a user**.

It should be noted that the *porthole*, which is a *subgraph of the Web*, generated by the *focused crawler* of Chakrabarti et al. is equivalent to the claimed **linkage structure** that is learned. It should further be noted that the generation of a porthole or specialized link structure (p 20, last paragraph) is equivalent to the claimed **learning a linkage structure**.

Chakrabarti et al. do not explicitly teach **collecting at least a set of aggregate statistical information and a set of predicate-specific statistical information**.

Chaudhuri et al. teach that the MNSA technique for determining if the existing set of statistics contains an essential set of statistics should be qualified as follows. First, note that even for a single selectivity variable, multiple statistics may be applicable with different degrees of accuracy. Second, for an SPJ query, MNSA guarantees inclusion of an essential set of the query only as long as the selectivity of predicates in the query is between g and 1-g. Third, although for SPJ queries MNSA ensures that an essential set is included among the statistics, it is necessary to extend the method beyond simple queries. Aggregation clauses can be handled by associating a selectivity variable that indicates the fraction of rows in the table with distinct values of the column(s) in the clause (Column 19, lines 35 – 63), which meet the limitation of **collecting at least a set**

of aggregate statistical information and a set of predicate-specific statistical information.

Because both Chakrabarti et al. and Chaudhuri et al. teach methods of collecting statistics, it would have been obvious to one skilled in the art to substitute one method for the other to achieve the predictable result of collecting aggregate and predicate-specific statistics.

Regarding dependent claim 2, Chakrabarti et al. teach that Query construction is not a one-time investment, because as pages on the topic are discovered, their additional vocabulary must be folded in manually into the query for continued discovery (p 7, lines 4 – 6), which meet the limitation of **the user-defined predicate specifies content associated with a document**. It should be noted that the *additional vocabulary of pages on the topic* of Chakrabati et al. is equivalent to the claimed **content associated with a document**.

Regarding dependent claims 3 and 4, Chakrabarti et al. teach that pages that are examples associated with a topic can be preprocessed as desired by the system. The user's interest is characterized by a subset of topics that is marked good. No good topic is an ancestor of another good topic. Ancestors of good topics are called path topics. Given a Web page, a measure of its relevance must be specified to the system (p 8, lines 9 – 14), which meet the limitation of **the statistical information collection step uses content of the one or more retrieved documents** and that **the statistical**

information collection step considers whether the user-defined predicate has been satisfied by the one or more retrieved documents, since a determination is made about the ancestors and preprocessed pages are used, which are equivalent to the claimed **one or more retrieved documents**. It should be noted that the *topic* of Chakrabarti et al. is equivalent to the claimed **content** and **predicate**.

Regarding dependent claims 5 and 6, Chakrabarti et al. teach that we have presented evidence in this section that focused crawling is capable of steadily collecting relevant resources and identifying popular, high-content sites from the crawl, as well as regions of high relevance, to guide itself. It is robust to different starting conditions, and finds good resources that are quite far from its starting point. In comparison, standard crawlers get quickly lost in the noise, even when starting from the same URLs (p 20, Section 4.8 and p 18, Figure 9), which meet the limitation of **the collected statistical information is used to direct further document retrieval operations toward documents which are similar to the one or more retrieved documents that also satisfy the predicate**, and that **the collected statistical information is used to direct further document retrieval operations toward documents which are more likely to satisfy the predicate than would otherwise occur with respect to document retrieval operations that are not directed using the collected statistical information**, since the focused crawling of Chakrabati et al. utilizes statistical information (p 3) and compares their crawler to other crawlers and outlines the other's shortcomings (Fig 9).

Regarding dependent claim 7, Chakrabarti et al. teach that multiple citations from a single document are likely to cite semantically related documents as well. This is why the distiller is used to identify pages with large numbers of links to relevant pages (p 8, last paragraph), which meet the limitation of **the collected statistical information is used to direct further document retrieval operations toward documents which are linked to by other documents which also satisfy the predicate**. It should be noted that the semantically related documents of Chakrabarti et al. is equivalent to the claimed **documents which are linked to by other documents which also satisfy the predicate**

Regarding dependent claim 8, Chakrabarti et al. teach that we describe a Focused Crawler, which seeks, acquires, indexes, and maintains pages on a specific set of topics that represent a relatively narrow segment of the Web. Thus, Web content can be managed by a distributed team of focused crawlers, each specializing in one or a few topics (p 2, fourth paragraph), which meet the limitation of **the information network is the World Wide Web and a document is a web page**.

Regarding claims 10 – 17 and 19 – 26, the claims incorporate substantially similar subject matter as claims 1 – 8, and are rejected along the same rationale.

Claims 9, 18 and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chakrabarti et al. and Chaudhuri et al. (US 6529901 B1) as applied to claims 1 – 8,

10 – 17 and 19 – 26 above, and further in view of Chakrabarti et al. (Distributed Hypertext Resource Discovery Through Examples) [as cited by appellant] later referenced as Ch2 et al.

Regarding dependent claim 9, Chakrabati et al. do not explicitly teach that **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages**.

Ch2 et al. teach that other strategies are also known, such as, if the URL is of the form http://host /path, then the crawler may truncate components of path and try to fetch these URL's. If links could be traversed backward, e.g. using metadata at the server, the crawler may also fetch pages that point to the page being 'expanded.' (p 382, Column 1, lines 29 – 37), which meet the limitation of **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages**.

It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the teachings of Chakrabarti et al. and Chaudhuri et al. with that of Ch2 et al. because such a combination would provide the users of Chakrabarti et al. and Chaudhuri et al. with teachings of *the architecture of a hypertext resource discovery system using a relational database* (p 375, Column 1, lines 1 & 2).

Regarding claims 18 and 27, the claims incorporate substantially similar subject matter as claim 9, and are rejected along the same rationale.

(10) Response to Argument

Appellant argues that Chaudhuri does not teach the step of collecting at least a set of aggregate statistical information and a set of predicate-specific statistical information because Chaudhuri is directed toward a technique, MNSA, for determining if the existing set of statistics contains an essential set of statistics (p7).

The Office disagrees.

First, it should be noted that appellant initially argues that the statistics in Chaudhuri et al. is an existing set of statistics (p 7, second paragraph). Obviously, the skilled artisan is well aware that in order to have an existing set of statistics the statistics had to be collected at some point in time. The existence of a set of statistics only bolsters the Office's position.

Specifically, Chaudhuri et al. teach that note that even for a single selectivity variable, multiple statistics may be applicable with different degrees of accuracy. Second, for an SPJ query, MNSA guarantees inclusion of an essential set of the query only as long as the selectivity of predicates in the query is between g and 1-g. Third, although for SPJ queries MNSA ensures that an essential set is included among the statistics, it is necessary to extend the method beyond simple queries. Aggregation clauses can be handled by associating a selectivity variable that indicates the fraction of rows in the table with distinct values of the column(s) in the clause (Column 19, lines 35 - 63), which meet the limitation of collecting at least a set of aggregate statistical information and a set of predicate-specific statistical information.

Essentially, Chaudhuri et al. teach gathering statistics by handling aggregation clauses which is equivalent to the claimed set of aggregate statistical information. The appellant assets that the handling of the group by and/or select distinct clauses fails to teach the collecting of a set of information maintained for all documents (p 7, third paragraph) with no explanation as to why they are different. It should be noted that the appellant is confusing and complicating the issue.

The claim recites collecting a set of aggregate statistical information. The specification describes but does not define the aggregate statistical information to possibly be a set of information maintained for all documents by way of example. The group by and/or select distinct clauses of Chaudhuri et al. are examples of aggregation clauses discussed by Chaudhuri et al. The appellant appears to compare specific examples described in his specification with specific examples discussed in Chaudhuri et al. thus muddying the issue. Either aggregation clauses, which are queries that produce a set of statistics, or the existing set of statistics described by Chaudhuri et al. clearly meet the claimed limitation of collecting a set of aggregate statistical information.

In other words, if one chooses to limit the claimed set of aggregate statistical information to be a set of information maintained for all documents then the existing set of statistics described by Chaudhuri et al. meets that claimed language. However, if one chooses to rely upon the broadest, reasonable interpretation in light of the specification then the aggregation clauses, which are queries that produce a set of statistics described by Chaudhuri et al., meet the claimed language in question.

Likewise, Chaudhuri et al. further teach that an essential set of predicates from the SPJ queries are included among the statistics, which meet the claimed predicate-specific statistical information. Appellant again complicates the issue by focusing on a range. It should be noted that even if the teachings of Chaudhuri et al. place a range on the predicates in the query, the teachings still met the claim language of a set of predicate-specific statistical information within the broadest, reasonable interpretation in light of the specification. Appellant argues that an example of a set of predicate-specific statistical information described in the specification is information maintained for the subset of the retrieved documents which satisfy a given predicate. The range disclosed in Chaudhuri et al. is clearly a subset.

Appellant argues the motivation to combine the references (p 8).

The office disagrees.

First, it should be noted that while the Court quoting In re Kahn, 441 F.3d 977, 988, 78 USPQ2d 1329, 1336 (Fed. Cir. 2006), stated that "'[R]ejections on obviousness cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.'" KSR, 550 U.S. at ___, 82 USPQ2d at 1396. Exemplary rationales that may support a conclusion of obviousness include, among others, simple substitution of one known element for another to obtain predictable results (MPEP 2141 III).

Specifically, The Office maintains that because both Chakrabarti et al. and Chaudhuri et al. teach methods of collecting statistics, it would have been obvious to

one skilled in the art to substitute one method for the other to achieve the predictable result of collecting aggregate and predicate-specific statistics.

It is not understood how, in the case of a claim to a combination, that one of ordinary skill in the art could not have combined the claimed elements by known methods (such as technological difficulties); the elements in combination do not merely perform the function that each element performs separately; or the results of the claimed combination were unexpected.

In response to appellant's argument that the examiner's conclusion of obviousness is based upon improper hindsight reasoning, it must be recognized that any judgment on obviousness is in a sense necessarily a reconstruction based upon hindsight reasoning. But so long as it takes into account only knowledge which was within the level of ordinary skill at the time the claimed invention was made, and does not include knowledge gleaned only from the appellant's disclosure, such a reconstruction is proper. See *In re McLaughlin*, 443 F.2d 1392, 170 USPQ 209 (CCPA 1971).

Appellant argues that Ch2 et al. fail to teach that **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages** because the tokens are not used in statistical collection (p 9 and 10).

The Office disagrees.

First, it should be noted that appellant is not only performing a piecemeal analysis by analyzing the teachings of Ch2 et al. in a bubble and not as being apart of a combination of teachings but appellant also takes the cited portion of the reference out of context as well. It bears repeating that a reference is valid for all that it teaches (MPEP 2123).

The cited portions of Ch2 et al. are within a larger context than appellant gives credence. Ch2 et al. teach at the beginning of the section that now we will describe how the scores determined by the classifier and distiller are combined with other ***per- URL*** and per-server ***statistics*** to guide the crawler. To make the discussion concrete, we give a specific design, but it is important to note the flexibility of the architecture to supporting other policies and designs as well (p 381, column 2, last paragraph).

The teachings of per-URL statistics by truncating components of URLs clearly meet the claim language of gathering statistical information by using URL tokens.

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,

/Nathan Hillery/

Primary Examiner, Art Unit 2176

Conferees:

/Doug Hutton/
Doug Hutton
Supervisory Primary Examiner
Technology Center 2100

/Rachna S Desai/
Primary Examiner, Art Unit 2176